The Report Committee for Yuanhan Li
Certifies that this is the approved version of the following report:

HANDLING COMPLEX MULTILEVEL DATA STRUCTURES

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

____
S.Natasha Beretvas

____
Tse-min Lin

HANDLING COMPLEX MULTILEVEL DATA STRUCTURES

by

Yuanhan Li, B.Eco.

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Statistics

The University of Texas at Austin

May 2013

Abstract

HANDLING COMPLEX MULTILEVEL DATA STRUCTURES

by

Yuanhan Li, M.S. Stat

The University of Texas at Austin, 2013

SUPERVISOR: S.Natasha Beretvas

This report focuses on introducing two statistical models for dealing with data involving complex social structures. Appropriate handling of data structures is a concern in the context of educational settings. From base single-level data to complex hierarchical with cross-classifications and multiple-memberships, we explain and demonstrate their distinction and establish appropriate regression models. Real data from the National Center for Education Statistics (NECS) is used to demonstrate different way of handling a cross-classified data structure as well as appropriate models. Results will be presented and compared to examine the practical operation for each model.

Keywords: Hierarchical Model, Cross-Classification, Multiple-Membership

Table of Contents

List of Tables

List of Figures

Chapter One

Introduction

The motivation of this report is partly to introduce some key ideas of fairly recent statistical methodology for dealing with data on complex social structures in an understandable and readable way. This report focuses on depicting the hierarchical data structures as well as the factors which may give rise to the cross-classifications, and formulating the relevant models.

There are several text books concerning multilevel and cross-classified structures and models (Snijder & Bosker, 1999; Raudenbush & Bryk, 2002; Hox, 2002; Beretvas, 2008). A couple of books also provide the introduction of both cross-classification and multiple membership models (Goldstein, 2003; Rasbash & Browne, 2001). Fielding and Goldstein (2006) offers several examples regarding complex multilevel structure modeling.

Since many of the key ideas of statistical modeling and statistical control of variables are elaborated well in traditional explanatory regression, at Chapter 2, we still consider a brief review of the simple regression model to establish notations which is essential to understand as the statistical models become more complex. The more complex situation arises thereafter including hierarchical structures with example of pupils within classes nested in schools, and why standard regression models should be extended to Hierarchical Linear Models.

Chapter 3 then involves an extension of base multilevel models using an educational example. We allow the pupils of secondary schools have a parallel hierarchy across primary schools. The same example is then extended to introduce when the Cross-

Classification Random Effect Models should be used as an appropriate way of handling data in such structures.

An extension of Cross-Classification structure is introduced in Chapter 4. The idea of Multiple-Membership is applied in further complexity where pupil's mobility is incorporated. A strict hierarchy of pupils within institutions is no longer applicable when pupils attend more than one institution during a period of time. The corresponding Multiple-Membership Random Effect Model is introduced to show these institution random effects as weighted contribution.

Chapter 5 then summarizes three type of structures discussed in previous sections as a more general form. Here, classification diagrams and general notation are included to help the reader identify the main differences and correlations of HLM, CCREM and MMREM.

The last chapter then entails a real data analysis where educational data from National Center for Education Statistics (NECS) is used to reveal the relationship of pupils' academic achievements and their learning experience from attended schools. In this educational setting, a set of pupils may be crossed with a set of schools in each time of points. In terms of the multiple possibilities of data structure analyses, three models will be considered to accommodate corresponding structure. Results will be presented and compared to examine the practical operation for each model.

Chapter Two

Basic HLM in Hierarchical Social Structures

2.1 Explanatory Models Using Simple Regression

The aim of many statistical models is to try to analyze the variation in the targeted response variable which is explained by a set of one or more explanatory variables. For the response variable, in this report, we will focus on continuous response variables, which are commonly used and which are typically modeled under an assumption of normality. In addition, this study is intended to focus on consideration of the data's structure. Particular importance of complex structures involving hierarchies of observation units will be the focus. Hence, we will introduce the simple regression model and then extend to accommodate more complex structures, like the conventional multilevel model and its extensions.

Some main, relevant ideas about conventional linear regression will be presented here through an example. Suppose we are interested in the relationship between pupils' Reading Standard Test Score at the end of their 4th semester (after 2 kindergarten semesters and 1 primary school semester) and the random effect of the schools that they attended. We may also believe that it might be fruitful to explore, for instance, differences due to pupils' genders and family socio-economic status, etc. Since gender and family socioeconomic status may also affect students' aptitude before attending schools, these controlling variables might have prior effects except for school influences. At this stage we assume we have no information on the schools attended by sample data on pupils. Thus we denote y_i as the response observation for a particular pupil *i*, and correspondingly x_{1i} as the gender variable. This gender variable is a binary variable taking only two values and treated by "dummy" indicators. Hence, we have $x_{1i} = 1$ if the pupil is male and 0 otherwise. The usual simple regression model for this situation is

$$y_i = \beta_0 + \beta_1 x_{1i} + e_i \tag{1}$$

Here x_{1i} are observed as values of "explanatory variable" and y_i is the response or dependent variable observation. The coefficients β_0, β_1 are called parameters of the model and will be estimated in some way from the data.

The initial interest is in the coefficient β_1, the effect of gender on C4RSCL score (a standard reading test score). Considering there is only one explanatory variable included in the model, its interpretation is the average effect of gender, that indicating a difference of one unit on the male from female average score will yield an expected difference of β_1 in the response. The quantity β_0 is known as "intercept", interpreted as the average of y when all the explanatory variables are 0. From the way the variables have been defined we see that the intercept β_0 is the average C4RSCL score (y) for female pupils ($x_{1i} = 0$). In this situation, the model is specified $y_i = \beta_0 + e_i$. Therefore, female pupils are often referred to as the base or reference category. In regression models, this kind of coefficients of variables that have been explicitly included as explanatory in the model are often called "fixed effects", and the fixed part in our example, $\beta_0 + \beta_1 x_{1i}$ is often calculated to be a predictor of y.

When it comes to the term e_i, it is usually known as the "disturbance", "residual" or "error" and it is pivotal to analyses when we use statistical models. This term is included because we can never hope to fit all the empirical data exactly by a model no matter how complex that model might be. Thus, we hope the model is a reasonable approximation to reality and this residual indicated the extent to which the deviation of actual observed value from the fixed part prediction of y in its model. In the present context, the disturbance term represent the unobserved variability in pupils' C4RSCL responses due to effects that we either cannot observe or did not include in the model. Apart from the gender, it is possible that there are some other influential characteristics. These factors vary between pupils and may give rise to the actual responses higher or lower than the expected value obtained in the regression models given the explanatory variable values. An appropriate model is one whose unobserved variability does not impact the fitness in a system way. In other words, in our example, at given level of explanatory variable, the residual effect on all the pupils' C4RSCL average out to 0. The scale of all these unobserved and excluded possible combined sources reflected in e_i is indicated by the extent of the variability of e_i, which is measured by its variance, denoted as σ_e^2. σ_e^2 is another parameter of a model which can be estimated from the data and also a basis for measures of Goodness of Fit like other fixed effects parameters.

There are some assumptions required to ensure certain desirable statistical properties in the model. First, a common assumption is that each e_i is normally distributed with mean 0 and variance σ_e^2. Thus, the residual random effects should have expectation 0; Secondly, they should be uncorrelated with explanatory variables in the model. Moreover, whatever the observation *i*, the observation of e_i should be constant, that is, we do not expect the variability of the random effect in an observation to be influenced by any particular characteristic of the observation, like any of the explanatory fixed effects. Hence, the Ordinary Least Squares (OLS) estimation could be used in traditional regression models under these assumptions. Despite any influential factors, the pupils' reading scores are scattered and normally distributed around the average level, known as Unconditional Regression. When gender effect on pupil's reading performance is involved, all the pupils are divided into two groups, male (coded as 1) and female (coded as 0), with same variance in both groups. This is the representation of constancy of system variance σ_e^2, and also known as Conditional Regression model corresponding

Equation (1). By doing regression on Gender, we can reduce the uncertainty (random effect) in Unconditional Model to some extent.

2.2 Hierarchical Data Structures and HLM

2.2.1 Hierarchical Data Structure

The framework in the previous section may be too simplistic since almost all kinds of social and educational data have hierarchical or clustered structures. Take our previous context for instance, each pupil (level-1) studies in classes (level-2) which in turn are nested within secondary schools, so this is a three-level hierarchy. Here we use tables and figures to help illustrate the data's structure.

Figure1 and Figure2 show the affiliations of lowest level units (pupils) with higher level units (classes and schools). Since there is no crossing of lines, the data's structure can be regarded as a Pure Hierarchy. Consideration of clustering pupils into classes within schools is due to the reason that shared education circumstances (classes, teachers and schools) influences children from the same school may tend to be more like each other than pupils chosen from the population. To be more specific, we keep taking advantage of same context and extend it further by adding the clustering consideration into Figure 3. The points surrounded by dotted ellipse are nested in school 1. In terms of male pupils, the average reading score of male pupils in School1 is higher than School2. If School1 is a particularly effective school, it is reasonable that most children may have a better academic results than expected from knowing just their individual characteristics (such as gender, we discussed above). As we pointed out in previous section, the simple regression based on no information of schools where we sampled pupils nested in, we may focus too much on individual factors and ignore the location of the children in these shared contexts. This is the shortcoming of simple regression comparing to the hierarchical linear model.

In purely hierarchical structure, the lowest level unit is affiliated with a single higher level clustering unit although there might be multiple lower level units per higher level cluster. Come back to our example, as shown in Figure1, we can easily determine one pupil's school (higher cluster) if only given his/her class (the lowest cluster). If we know the class ID, we will know the corresponding school because class has ONLY ONE link to one school, which leads to no cross line appearing in network graph. We can also use a table to help identify the clusters. In Table1, the columns represent the lower level cluster (class) and the rows represent the higher level cluster (school). If the data are pure hierarchy, then the lower cluster (column) will has only one affiliation to a certain higher cluster (row), in our example, exactly one cell per column and vice versa.

2.2.2 Basic Hierarchical Linear Models

We wish to model a relationship between the individual response and the explanatory variables, taking into account the possibility that this relationship may vary across the clusters where the individuals are nested in. Raudenbush and Bryk's (2002) have discussed the conventional Hierarchical Linear Model (HLM) and we could take advantage of their levels formula to our example. In basic HLM model, we consider the lowest level units are purely clustered in higher level classifications. In the present case, we shall refer to the classes and schools as higher level units and pupils as lower level units. For simplicity's sake, we just begin with school effect on pupils' scores (see Figure3) to demonstrate the frame of model and then extend to more complex case (with classes effects). Thus, we have schools as level-2 units, and pupils as level-1 units. A simple such model can be written as follows:

At level one:

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + e_{ij} \tag{2}$$

At level two:

$$\beta_{0j} = \beta_{00} + u_{0j} \tag{3}$$

$$e_{ij} \sim N(0, \sigma^2_{e_0})$$

$$u_{0j} \sim N(0, \sigma^2_{u_0})$$

Where Y_{ij} is reading test scores for pupil i who attended school j, X_{ij} is predictable variable (here, is pupil's gender), residual term e_{ij} explains the lowest level deviation for each pupil, and u_{0j} explains the school level deviation. They are assumed independently and normally distributed with mean of zero and variances of $\sigma^2_{e_0}$ and $\sigma^2_{u_0}$ respectively. Combining Equation 2 and 3, we get the single equation:

$$Y_{ij} = \beta_{00} + \beta_{1j} X_{ij} + u_{0j} + e_{ij} \tag{4}$$

Note that in this conditional model, the effect of the level-1 predictor is assumed as fixed across schools in equation 4, and the net effect of differences between two genders is the same within all schools as evidence by the fact that the lines are parallel and hence have the same slop (see Figure 3, the constant coefficient of X is depicted as parallel lines with same slopes for school 1 and school 2). It shows a situation where on average, both male and female pupils in certain schools are higher or lower than those in other schools by fixed amount.

This model could be viewed as a standard analysis of covariance if we treated each u_{0j} as a fixed parameter to be estimated. Such a model however will often be inappropriate, for the following reasons.

First, if we consider the school effect u_{0j} as fixed part effects, one way of representing this sort of school effect is a standard multiple regression model as we previously discussed, thus we should use a set of dummy indicator variables for schools, with one less than the number of schools. One school will be arbitrarily chosen to become the baseline against which other schools are referenced. In some circumstances where we have just a few schools, this might be a reasonable approach. But for large number of schools, this method will be too tedious leading too many parameters to be estimated. Second, some of schools may have very few individuals, so that their individual departures will be poorly estimated. Most importantly, we may be interested in treating the schools as sample from a population of schools and with to make a general inference about the likely behavior of other schools in this population rather than, estimating separately for each school in sample. For all these reasons, it will usually be more appropriate to regard u_{0j} as random variable representing random deviation from the distribution of school effect at school level in the same way as we might regard the e_{ij} terms as a random unobserved effects operating at the pupil level.

Such random effects can be illustrated in Figure 3, where we can see, for pupil A (red spade dot) who attend school1, his total deviation from estimated value $\hat{y}_{ASchoo1}$ (on solid regression line) can be decomposed to two parts, the deviation from pupil A to school1 average line (upper dashed line) and the school1's deviation from fixed part (estimated value). In conventional HLM, we merely measure the random effect for each pupil at an overall level. Hence, the model now has two random variables specifying two sources of variation, at the level-1 of pupils e_{ij} and at level-2 of schools u_{0j}. Such model is also often known as "variance components models". The two components, $\sigma_{e_0}^2$ and $\sigma_{u_0}^2$ will also need to be estimated along with β_{0j}, β_{1j}.

Therefore, the total residual variance is $\sigma_{e_0}^2 + \sigma_{u_0}^2$, and we usually interest in $\frac{\sigma_{u_0}^2}{\sigma_{u_0}^2 + \sigma_{e_0}^2}$, the proportion of the total variance that is attributable to schools, considered as a summary of significance of school effects. It can be shown that the size of correlation between the pupils who are clustered in same school, often know as Intra-Class Correlation (ICC). This is one reason why traditional OLS is often inappropriate, since in this multilevel situation, traditional OLS treats the total residual term in the model as $u_{0j} + e_{ij}$, only a single random variable. Whereas, our HLM makes two random variables at each level of data structure, which can explicitly recognize the resources of variations.

Chapter Three

Cross-Classification Data Structures

3.1 The Nature of Cross-Classification

It is commonly the case that data have a structure which is not purely hierarchical. Subjects may be clustered not only into hierarchically ordered units (e.g., in our previous case, classes are nested within schools, leading to a strict hierarchy for each pupil), but we may also belong to more than one type of unit at a given level of a hierarchy. This led to a desire to extend the methodology to handle the analysis of effects in such structures. The separation of effects arising from what we may regard as random cross-classification in one such extension.

In a basic two-level multilevel structure, pupils may be classified hierarchically by their school. In many cases, however, such units may be classified in more than one dimension. For example, a pupil might be clustered sequentially to a particular combination of primary school and secondary school. The essential difference between the cross-classification structure and basic pure hierarchy is at the affiliation of the classifications. Only if primary schools belong to certain secondary school, these structures could be considered as three-level (including pupil level) conventional hierarchies. Otherwise, when primary schools and secondary schools are less dependent, these structures are still a two-level model in this sense, but the level-2 units are now combination of particular primary schools and secondary schools. Here, we illuminate the cross-classification case with the help of network graph in Figure4.

The crossed lines in the network graph (Figure4 and Figure5) clearly infer that the pupils are cross-classified by primary school and high school. The difference between Figure4 and Figure5 is the classifying order. Such diagrams, however, can become far too elaborate for some of the more complex structures and might become more confusing. Thus, we may want to use alternative pictorial representations, see Table 2. Table representation can also help to infer the cross-classification, because neither the row nor the column has single cell with elements, which could give rise to cross lines in network graph.

3.2 Cross-Classification Random Effect Model

3.2.1 Objective of Cross-Classification Random Effect Model

The traditional two-level HLM, as we discussed in previous section, takes into account only two parts of random effects, the secondary schools' effect and pupils' own effect. Pupils coming from same primary school may also vary significantly. Assume some students' reading scores are lower than the others', this may be due in part to the fact that this secondary school drew these students from a certain low efficient primary school which is influential to student outcomes. However, in our conventional HLM, this primary school impact will likely inappropriately be counted as part of individual random effect (e_{ij}) because the HLM fails to explain partition variability also into a primary school effect.

Therefore, a model designed to handle cross-classification structures is essential. If we ignored the fact that some secondary school effects in addition to primary school effects on the pupil outcome, we could be dealing with what is often called an "underspecified model", meaning the estimation of other effects might be poor. The object of cross-classification model is to give us an idea of the extent to which variation in outcomes might be attributable to unobserved influences at the level of each of the three types of units (Secondary school units, primary school units and pupil units) in the model.

3.2.2 Notation for Cross-Classified Random Effect Model

In a basic two-level model we denoted a level-2 random intercept effect for level-2 unit j by u_{0j} and its variance by $\sigma^2_{u_0}$. Level-2 effects are now more complex in the Cross-Classified Random Effect Model (CCREM) and we require extending the notation. Firstly, the level-2 effect cell is combination of two factors. If we use j_1 to indicate a particular unit of the first classification and j_2 for second factor, then a particular level-2 unit now becomes (j_1, j_2). Correspondingly, the level-2 effect is the sum of two separate random effects which we now denote as $\{u^{(1)}_{0j_1} + u^{(2)}_{0j_2}\}$. Using a similar formulation as before, with the cell notation, the model for level-1 unit i within the level-2 unit (j_1, j_2) can now be written as

At level one:

$$Y_{i(j_1,j_2)} = \beta_{0(j_1,j_2)} + X_{i(j_1,j_2)}\beta_{1(j_1,j_2)} + e_{i(j_1,j_2)} \quad (5)$$

At level two:

$$\beta_{0(j_1,j_2)} = \beta_{00} + u^{(1)}_{0j_1} + u^{(2)}_{0j_2} \quad (6)$$

$$e_{i(j_1,j_2)} \sim N(0, \sigma^2_{e_0})$$

$$u^{(1)}_{0j_1} \sim N(0, \sigma^2_{u^{(1)}_0})$$

$$u^{(2)}_{0j_1} \sim N(0, \sigma^2_{u^{(2)}_0})$$

In our pupil reading test example, $\beta_{0(j_1,j_2)}$ is intercept term for each cross-classification combination units, which can be decomposed into three parts (see Equation 6), the overall mean β_{00}, the primary school random effect term $u^{(1)}_{0j_1}$ and the secondary school random effect term $u^{(2)}_{0j_2}$. β_{00} is the fixed and predicted outcome score averaged across primary school and secondary school. Here we denote the higher level of the classification unit, the further to the right will appear the associated subscript. Since we have two-level structure, $u^{(1)}_{0j_1}$ and $u^{(2)}_{0j_2}$ can both represent the level two random effect, we therefore add a superscript to distinguish two classifications at same level. $X_{i(j_1,j_2)}$ is pupil-level descriptor(here, the pupil gender), to explain variability in the outcome. The coefficient $\beta_{1(j_1,j_2)}$ is assumed to be fixed across the primary school and secondary school. We still refer to basic level-1 variance by $\sigma^2_{e_0}$ whereas the level-2 variance is now $\sigma^2_{u_0^{(1)}}$ $+\sigma^2_{u_0^{(2)}}$, the sum of two additive components corresponding to each factor in the cross-classification.

Thus, the single equation with combination of two levels is:

$$Y_{i(j_1,j_2)} = \beta_{00} + X_{i(j_1,j_2)}\beta_{1(j_1,j_2)} + u^{(1)}_{0j_1} + u^{(2)}_{0j_2} + e_{i(j_1,j_2)} \tag{7}$$

There is one more point we should pay attention to. It is also possible that the cross-classification variation may incorporate an "interaction term" denoted by $u^{(3)}_{0(j_1,j_2)}$. This will make the additive contribution of the random effect and in this situation, the cell effect is now characterized as the sum of three additive components $\{u^{(1)}_{0j_1} + u^{(2)}_{0j_2} + u^{(3)}_{0(j_1,j_2)}\}$. The motivation to consider the interaction term can be explained by our example. In some cases, the effect of primary school j_1 may not be independent of the effect of secondary school j_2. That means, it is possible that the marginal effect of primary school might differ according to which secondary school the pupil went to. Hence, the corresponding cross-classification level variance is the sum of three components, $\{\sigma^2_{u_0^{(1)}} + \sigma^2_{u_0^{(2)}} + \sigma^2_{u_0^{(3)}}\}$, where $\sigma^2_{u_0^{(3)}}$ is variance of interaction effect called interaction variance. However, it is hard to separate the level-1 (pupil) random effect $\sigma^2_{e_0}$ from the interaction effect $\sigma^2_{u_0^{(3)}}$ unless the all the cell sample sizes are large enough (Raudenbush&Bryk, 2002). So we usually set interaction term to zero in CCREM.

Like the purely clustered multilevel models, we can also use intra-class correlation to quantify the contribution of clustering classification to the total variance. For instance, the variance component of primary school can be calculated as

$\frac{\sigma^2_{u_0^{(1)}}}{\sigma^2_{u_0^{(1)}} + \sigma^2_{u_0^{(2)}} + \sigma^2_{e_0}}$, which represent the correlation between the pupil reading test scores in same secondary school but from different primary schools. The ICC can be used to explain the proportion of variability attributable to each clustering factors.

Chapter Four

Multiple-Membership Data Structures

4.1 The Nature of Multiple-Membership Data Structure

Multiple-Membership structure is a special case of Cross-Classification which is also a complex structure where we wish to disentangle its effect. Multiple-membership model might appear in situations where lowest level unit might be members of more than one of the same group at a higher level in the hierarchy. A student might be classified as belonging sequentially to a particular combination of primary school and secondary schools. Alternatively, a particular student may attend in one secondary school and move to another secondary school. In our previous data structure, either pure hierarchy or cross-classification, one student can only attend one type of higher unit for one time, meaning one particular student is supposed to stay in the same primary school for the whole of the elementary stage. But in reality, many students may change their schools due to various reasons. Those mobile students have been exposed to the “effects” of more than one school.

We keep using the network graphs and table to assist in gaining an intuition about the structure of multiple-membership data structure (see Figure6). Here we consider the secondary schools effects on pupils reading test scores assuming pupils change schools during this period. We can notice the cross lines appear in Figure6 and it can be treated as cross-classification by this means. But the difference between Figure 4 and Figure 5 is the number of clustering. In Figure 4,5 there are two clustering groups, primary and secondary schools, and the less purely nested relationship between them give rise to the cross connections. Nevertheless, in Figure 6, we can also find the cross lines in one group of clustering, because the multiple dash lines coming from lowest level units instead generate the cross connections. In this example, pupil 2,7,10 has multiple memberships while others do not transfer during this period. Therefore, the resource of crossings can help determine the data structure, in addition there must be at least two clustering variables for cross-classification to occur whereas one for multiple membership.

4.2 Multiple-Membership Random Effect Models

Our original objective of multilevel models is to disentangle and control for the higher level effects of particular clustering factor. However, like educational data depicted in Figure 6, we may encounter a problem as to how we might model the secondary school effect for observations which is attributable to more than one of these schools. A common and basic necessity is to assume that there is a known weight for each higher level unit to a lower level unit to apply to the school effects. Usually the weights of one group of schools to individuals are added up to unity. The choice of weights is to some extent subjective. The simply and uninformative way is to assign equal weights. In terms of our example, it may also be reasonable to assign weights proportional to time spent in each school by pupils. Sometimes, it might be thought that more recent school experience has a greater impact might counteract to some extent the time experience.

4.3 Notation for Multiple-Membership Random Effect Model

Considering now just the secondary schools, suppose that we know for each individual, the weight π_{ij}, associated with the j^{th} secondary school attended by pupil i with $\sum_{j=1}^{J} \pi_{ij} = 1$. Note that we allow the possibility that for those pupils who are only enrolled to one school, thus the probabilities to their schools are one and the remainders are zero. Note that when all level-1 units have a single non-zero weight of 1, we obtain the usual purely hierarchical model. Following Hill and Goldstain (1998) and Rasbash and Browne (2001), the multiple-membership random effect model (MMREM) can be written as follows:

$$Y_{i\{j\}} = X_{i\{j\}}\beta_{i\{j\}} + \sum_{h=1}^{J} u_{0h}\pi_{ih} + e_{i\{j\}} \tag{8}$$

$$u_{0h} \sim N(0, \sigma^2_{u_0})$$

$$e_{i\{j\}} \sim N(0, \sigma^2_{e_0})$$

$$with \textstyle\sum_{h=1}^{J} \pi_{ih} = 1 \text{ for each level-1 unit (pupil)}$$

Note that $\{j\}$ now means the full set of school $\{1,2,3 \cdots J\}$ and it is included as a subscript in the various quantities to represent the two-level multiple memberships. X is explanatory variable categorized as gender. The level-1 unit is indexes uniquely by i and the index h membership schools for each pupil i .The π_{ih} are the pre-determined weights using criteria we just have discussed. Summation of π_{ih} is equal to 1. Take two mobile pupils (pupil 2 and pupil 7) for instance.

For pupil2, who has attended two schools, SS1 and SS2, with equal weight. Then the model expression becomes

$$y_{2\{j\}} = X_{2\{j\}}\beta_{2\{j\}} + 0.5u_{0,1} + 0.5u_{0,2} + e_{2\{j\}}$$

Here the weights associated to other schools are all set to zero. For pupil 7, who has attended two schools, SS2 and SS3, we pre-assigned weights as 0.3, 0.7. Then the model expression becomes

$$y_{7\{j\}} = X_{7\{j\}}\beta_{7\{j\}} + 0.3u_{0,2} + 0.7u_{0,3} + e_{7\{j\}}$$

Chapter Five

General Notation for Cross-Classification and Multiple-Membership Structures

The notation and formula introduced in previous sections were those introduced in the early development of multilevel methodology. In this section, we introduce a more simplified notation to describe multilevel models by Browne et al (2001). It could make structure features of models implicit meanwhile understood by association with "classification diagrams". Usage of schematic representation and general notation can give us a more comprehensive and further understanding of the nature of complex structures. Figure 7 is the basic type of classification diagrams.

In Figure 7, *A* represents 2-level hierarchical model; *B* represents 2-level cross-classification at level-2 with combination of primary and secondary school; *C* represents multiple membership model.

Such diagrams enable us to classify the nature of hierarchical or cross-classifications or multiple memberships and to see how to fit into the appropriate framework given empirical data. The boxes represent classifiers at various levels. The same level classifiers are all listed at same horizontal level. Lower level units are affiliated to the classification in arrow's direction. Parallel arrows in Graph C represent multiple membership relationship.

The newer model notation does not involve complicated multiple subscripts. The basic two level variance component HLM associated with *A* in Figure 7 is written as:

$$y_i = (X\beta)_i + u^{(2)}_{school(i)} + u^{(1)}_{student(i)} \tag{9}$$

$$school(i) \in (1,2,\cdots\cdots J)$$

$$student(i) \in (1,2,\cdots\cdots N)$$

$$u^{(2)}_{school(i)} \sim N(0,\sigma^2_{u(2)})$$

$$u^{(1)}_{student(i)} \sim N(0, \sigma^2_{u(1)})$$

Now we have two classifications, indicated by superscript, classification 1 is student and classification 2 is school. Symbol $\in$ means “belong to” or “an element of”, hence $school(i)$ is one of J level-2 schools. Normality is usually assumed as indicated by last two normal distributions.

Together with B in Figure 7, the basis CCREM can be written as:

$$y_i = X_i\beta + u^{(3)}_{primaryschool(i)} + u^{(2)}_{\sec ondaryschool(i)} + u^{(1)}_{student(i)} \tag{10}$$

$$primaryschool(i) \in (1,2,\cdots\cdots J_3)$$

$$\sec ondaryschool(i) \in (1,2,\cdots\cdots J_2)$$

$$student(i) \in (1,2,\cdots\cdots N)$$

$$u^{(3)}_{primaryschool(i)} \sim N(0, \sigma^2_{u(3)})$$
$$u^{(2)}_{\sec daryschool(i)} \sim N(0, \sigma^2_{u(2)})$$
$$u^{(2)}_{student(i)} \sim N(0, \sigma^2_{u(1)})$$

The complexity of CCREM compared to HLM is merely extended for every additional number of crossed or hierarchical sets of any units at any levels. Note that the equation themselves do not explicitly show the nesting or crossings. In modern estimation procedure Monte Carlo Markov Chain (MCMC) who commonly used to deal with multilevel data, does not require knowing the exact nesting structure. We only need to provide the unique identifiers in the data for each classification.

However, we interpret in association with diagram C of Figure 7 as MMREM can be written as

$$y_i = X_i\beta + \sum_{h \in school(i)} u^{(2)}_h \pi_{ih} + u^{(1)}_{student(i)} \tag{11}$$

$$
\begin{aligned}
&school(i) \subset (1,2,\cdots\cdots J)\\
&student(i) \in (1,2,\cdots\cdots N)\\
&u_h^{(2)} \sim N(0,\sigma^2_{u(2)}) \ \text{ for } \ h \in \{j\} = \{1,2,\cdots\cdots J\}\\
&u^{(1)}_{student(i)} \sim N(0,\sigma^2_{u(1)})
\end{aligned}
$$

Here, we substitute $\in$ for $\subset$ since in MMREM can assign more than one school from the whole set $\{1,2,\cdots\cdots J\}$. The symbol $\subset$ means "a subset of" and therefore school (i) can have more than one school from the total set of school. Note that there is only one higher classifier factor $u_h^{(2)}$ in MMREM and the entire membership school share a same superscript (2). Browne et al (2001) also consider how the diagrams and model specifications can be extended to more general models with more levels, combinations of hierarchical, cross-classified or multiple membership structures.

Chapter Six

Real Data Analyses and Estimation Methodology

In this section, we will take advantage of different multilevel model to fit our real educational data. This analysis is not designed to study the theory of school effects on pupil academic outcome; instead, it aims at data structure analysis and flexibility of multiple model choices under different structure analysis results.

6.1 Dataset Description

This report used a subset of a large-scale state-wide dataset from EDAT data library in National Center for Education Statistics (NECS). In our study, we select 4th grade reading achievement scores (C4R4SCL) of a subset of 16,170 4th grade students were used as outcome. In EDAT data library, each sampled pupil was recorded at 7 time points, denoted by C1~C7, where C1 represents kindergarten fall semester to C7 representing the 8th spring semester. Here we used the reading test at the end of C4, 1st spring semester, as our dependent variable. As well as the targeted semester, we are also interested in the effects of pupil's previous study experience on C4 reading test score. Therefore, we use C1 (1st kindergarten spring) and C2 (2nd kindergarten fall) and collect each pupil's enrolled school at each of 3 stages. The dataset consisted of 1742 mobile pupils (10.66%) who attended two schools during C1, C2 and C4. None of the students attended 3 different schools across three stages. No missing data exist in dataset. One level-1 (pupil) predictor (gender) is included in the conditional model. School ID for each pupil at each of three stages is given.

6.2 Dataset Structure Analyses and Multilevel Models

In terms of our dataset, there is no pupil who attended 3 schools and only 10.66% out of total transferred school during the three measurement occasions. Considering the proportion of mobile students in the whole sample students is comparatively small, some researchers simply remove the mobile students' data from the dataset being analyzed (e.g., Lee 2000 and McCoach, O'Connell, Reis& Levitt, 2006). The dataset will therefore become a purely clustered dataset after deletion. To treat our dataset as simple hierarchical structure is the most convenient way to model the school effect, although deletion will lose some information of deleted students and their effect on the school they attend, which may lead to mis-specify the model.

Since we should consider the mobility of parts of students, we could directly apply the multiple-membership structure. Note that in our dataset, the schools from C1 to C4 are thought to be one classifier factor for ease of modeling and presentation. Therefore, our dataset structure can build a simple two-level multiple-membership which is typical designed to handle mobile students.

In addition, instead of considering this multiple-membership data structure, a cross-classification data structure can also be assumed. Since our dataset has only few cases of mobility to build multiple-membership which cannot be neglected, we may consider expressing these cross-classification cases through cross-classification data structure. The idea of this alternative representation is from our dataset's special characteristics-the mobility of one pupil can only occur once a learning stage. Thus, each pupil is believed to stay in one school at least one full semester. Given such a prior condition, we can build a two-level hierarchy (including pupil as level-1), where school level is associated to three learning stages. Therefore, at each learning stage, the classifying units are actually the same set of schools which are assumed independent. Such a transformation perfectly meets the basic cross-classification structure generation process—each classifier is not affiliated to others. Figure 8 can assist to illustrate such cross-classification scheme; Figure 9 is the corresponding multiple-membership version so that reader can compare the frameworks of two types of structures.

As shown in Figure 8, p3 transferred from S1 to S2 during C4; p7, p8 transferred from S2 to S3 during C4, others stayed in one school across three semesters. Note that even through one pupil attend only one school through C1, C2, C4, this school in different semester is considered to be uncorrelated and independent. Thus, in our cross-classification structure, there are three different clustering factors which determine the pupils' C4 reading test scores sequentially.

So far, we have figured out three possible structures with respect to our dataset and then we could construct corresponding multilevel models. For deleted nested hierarchy, the model is represented as Equation 8 where J means total number of sampled schools. For multiple-membership case, we are able to use Equation 10 where we could assign unequal weights proportional to the number of semesters which mobile pupils spend at one school. For cross-classification case, we use Equation 9, but the subscripts need slightly modification as follows:

$$y_i = X_i\beta + u^{(4)}_{C4school(i)} + u^{(3)}_{C2school(i)} + u^{(2)}_{C1school(i)} + u^{(1)}_{student(i)}$$

6.3 Estimation Approaches and Software

In this report, we select a modern Bayesian estimation method, Monte Carlo Markov Chain (MCMC) approach to accomplish our model estimation. The actual technicalities and philosophy of this approach are quite complex and will not be elaborated, here. To get the parameter estimates of interest, we need to perform a many-dimensional integration of posterior distribution, which is the historical stumbling block of the Bayesian approach. MCMC methods however circumvent this problem as they do not calculate the exact form of the posterior distribution but instead produce simulated draws from it. Thereafter, all the estimation and analyses are based on the sampled posterior distribution.

The software here we used is one specialist multilevel software named MLwiN. This software has been developed for fitting large and complex models using both frequentist likelihood and MCMC approaches. In MLwiN, MCMC estimation procedures are recent but they are rapidly developing and now encompass possibilities for a wide range of complex model structures such as we have discusses. MCMC sampling regards the likelihood-based frequentist estimation (IGLS) as start value along with default prior distribution, iterates sufficient times of sampling and obtains the targeted distribution sample from which we can calculate estimation of interest.

In this report, we begin with IGLS estimation value, set fixed parameter prior distribution as default, and assign all the random parameter prior as Gamma (1, 10000). Suppose 200,000 iterations and burn-in 180,000 iterations.

6.4 Results and Discussion

In MLwiN, a diagnostic criterion is used known as the Deviance Information Criterion (DIC) derived by Spiegelhalter et al. (2002). The DIC diagnostic is simple to calculate from an MCMC run as it simply the summation of average deviation from the 20000 iterations ($\overline{D}$) and the effective number of parameters (p_D). DIC therefore can be used to compare models as it consists of the two terms that measure the "fit" and the "complexity" of a particular model. Hence, a more fit (lower average deviation $\overline{D}$) and simpler (with less number of parameters p_D) model will be preferable with lower DIC value.

Here in our comparison of three models, DIC diagnostic will be calculated, and the fixed parameter and random effects' variance component estimates with standard error (SE) will be compared. The results are shown in Table 3. Gender coefficient is around -3.8 for all the three models, indicating that compared to girls, boys' reading capability is averagely lower by 3.8. MMREM intercept estimate is higher than others, but SE of MMREM fixed part parameters are smaller. The school total effect (ICC) under the CCREM is 0.43697 whereas MMREM ICC was 0.2445 and HLM was 0.246. Use of CCREM for partitioning the effect of each of the semesters make it possible to compare their influence on pupil's reading outcome. From table 3, we can know that pupil reading capability is most attributable to the first semester C1, which means the beginning of exposure to reading class plays a pivotal role in pupil's future reading study. As to the DIC diagnostic, CCREM has a lower value than MMREM and HLM by more than 30 points indicating a substantial difference in other two model's fit. To sum up, CCREM is favored over MMREM and HLM in terms of DIC.

This report is concentrating on the nature of different multilevel structures, therefore, to identify a more appropriate and reasonable data frame might be more important than fitness of model. Fit of model in return is dominated by a right recognition of data structure. It is not surprising to find more than one potential structure to represent same problem, like pupil's mobility, either multiple-membership and cross-classification works for it. In this report's scenario, cross-classification structure provides a better fit than multiple- membership merely because our dataset can meet several assumptions of the former one.

Although CCREM is more efficient than others in our example, the usage of CCREM to fit multiple-membership has its own restrictions. If we relax some CCREM assumptions, the advantage of CCREM over MMREM may be challenged. Firstly, interaction issue. In our estimated model, the two- and three- way interaction effects between classification factors were omitted. Actually, these classification units are essentially identical or barely changed. For simplicity's sake, we assume the identical factor in different time points may yield different uncorrelated impact on pupils. However, if this assumption is relaxed, the missing interactions will inflate the estimates of the cross-classification factors' random effects' variance components. Secondly, mobility issue. As we mentioned at previous section, the key to transfer multiple-membership to cross-classification is assuming each pupil can only move once during one time period. Only in this way, can we establish corresponding number of classifier factors. Clearly, this is not always the case. MMREM is free from such restriction because MMREM just focus on mobility results rather than the middle procedures. In short, we need pay more attention to the characteristics of different data structure, where our more efficient and appropriate multilevel model could be established.

Tables

Table 1

Pure Three-Level Clustering of Pupils within Classes within Schools

School	***Class***			
	C1	*C2*	*C3*	*C4*
S1	*p1 p2 p3*	*p4 p5*		
S2			*p6 p7 p8*	
S3				*p9 p10*

Table 2

Pupils Cross Nested within Primary Schools and Secondary Schools

Secondary School	***Primary school***			
	PS1	*PS2*	*PS3*	*PS4*
SS1	*p1 p2*	*p4 p5*		
SS2	*p3*		*p6 p8*	*p9*
SS3			*p7*	*p10*

Table 3

Comparison of HLM, MMREM and CCREM Model Parameter (Standard Error) Estimation and DIC Values

	Estimating Models		
	HLM	MMREM	CCREM
Fixed effects parameter			
Intercept	78.325 (0.301)	87.955 (0.297)	78.902 (0.335)
Gender	-3.717 (0.329)	-3.817 (0.328)	-3.789 (0.334)
Random effects variance			
Level One $\sigma^2_{e_0}$	451.871 (5.181)	499.658 (5.212)	440.108 (5.130)
Level Two σ^2_u	147.890 (7.887)	161.768 (8.844)	102.865^{C4} (7.087)
			109.127^{C2} (8.133)
			115.943^{C1} (8,733)
DIC	145543.43	145539.750	***145509.822***
p_D	845	921	1239

Note that, the HLM model estimated the effect of school at C4 semester. In CCREM, the random effects variance 102.865C4, 109.127C2, 115.943C1 refer to the school's random effect at C4, C2 and C1semester. The lowest DIC value is highlighted.

Figures

Figure 1: Network Graph of Pure Three-Level Clustering of Students within Classes within Schools

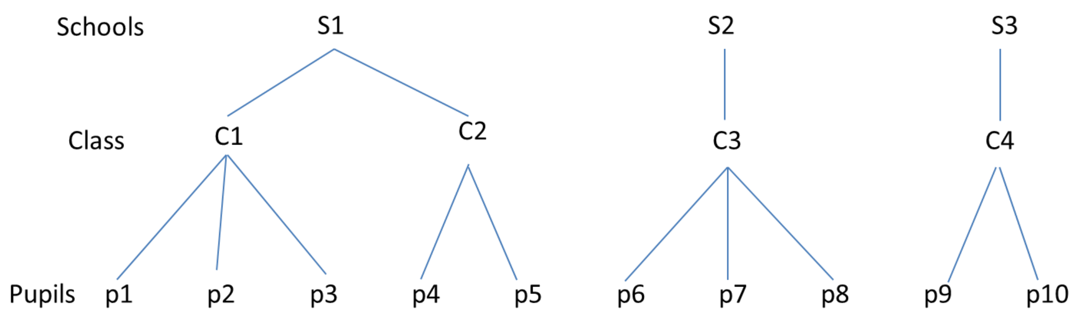


Figure 2: Alternative Representation of Figure 1 structure

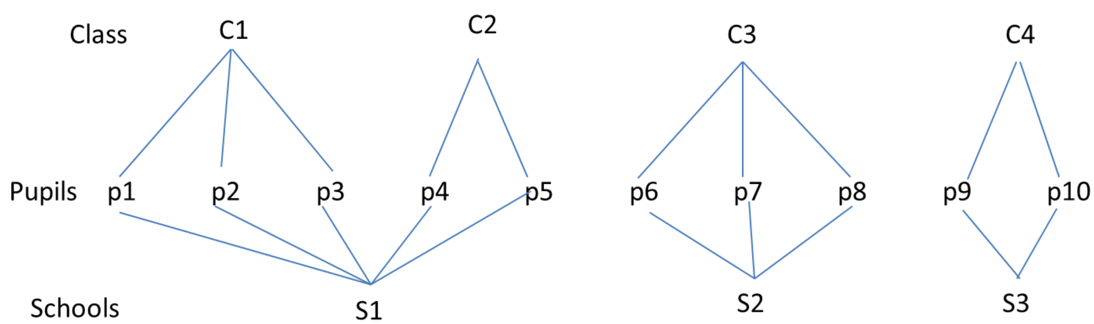
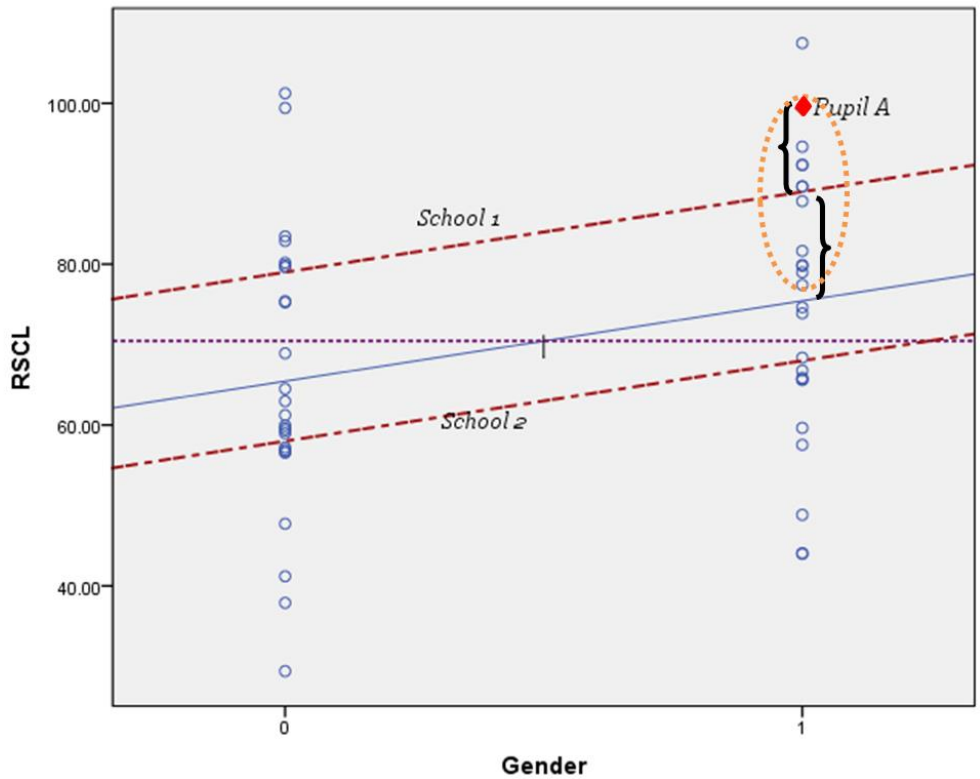


Figure 3: Multilevel Linear Regression Model with Pupil nested in Schools



Note. Solid line in middle of graph is the regression line. The points circled by orange dotted line represent the pupils who attended School1. School1 average outcome is marked as a parallel dashed line with Regression line. Red spade point is Pupil A. Distinction between A to dashed line is deviation within School (e_i) and distance from dashed line to regression line is deviation among schools (u_{0j})

Figure 4: Network Graph of Cross-Classification of Pupils by Primary Schools and Secondary Schools with Pupils listed by Primary Schools

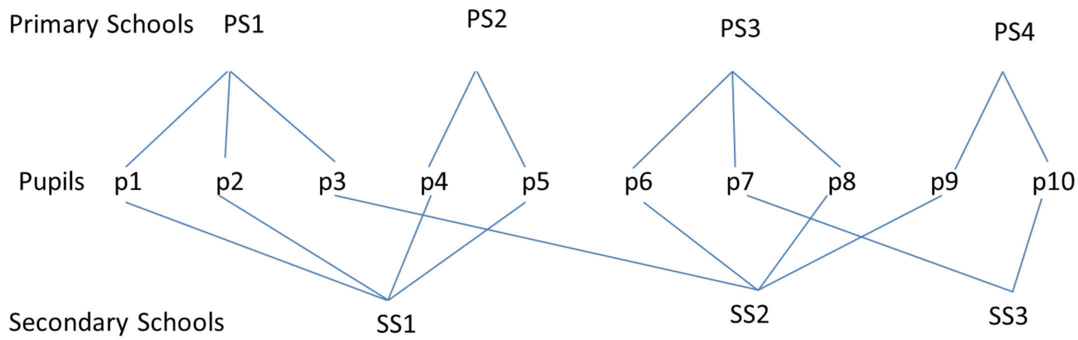


Figure 5: Network Graph of Cross-Classification of Pupils by Primary Schools and Secondary Schools with Pupils listed by Secondary Schools

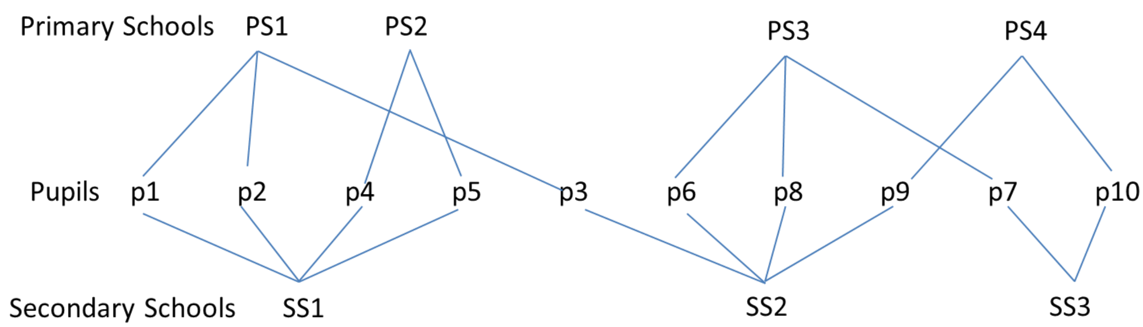


Figure 6: Network Graph Depicting Pupils attending Multiple Secondary Schools

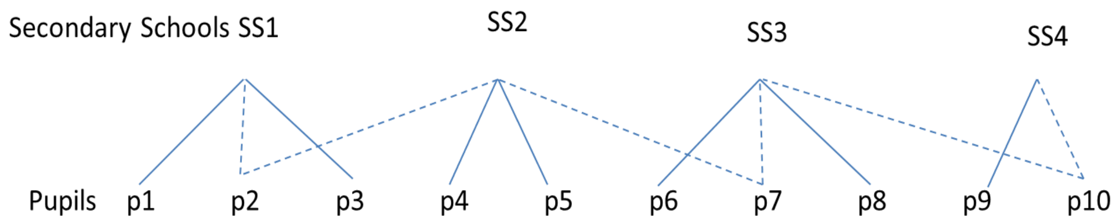


Figure 7: Classification Diagrams for Hierarchical, Cross-Classified and Multiple Membership Structures

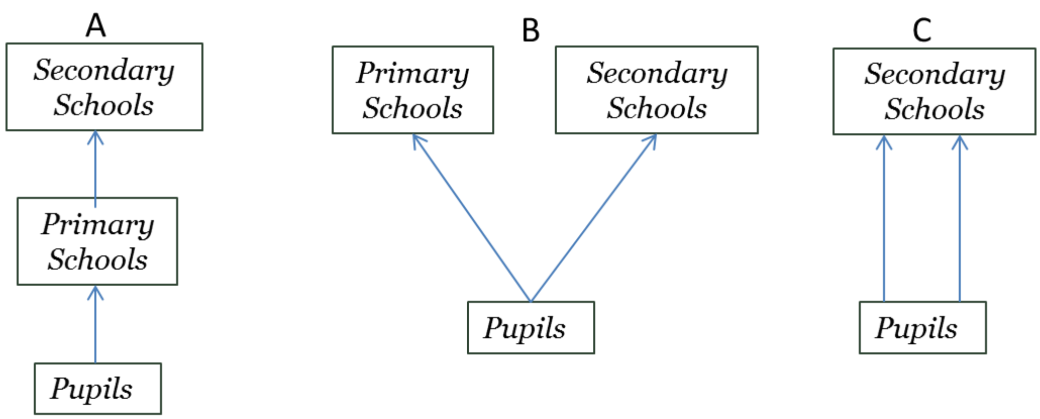


Figure 8: Network Graph illustrating Cross-Classified Structure with Pupils' Mobility at Each Semester

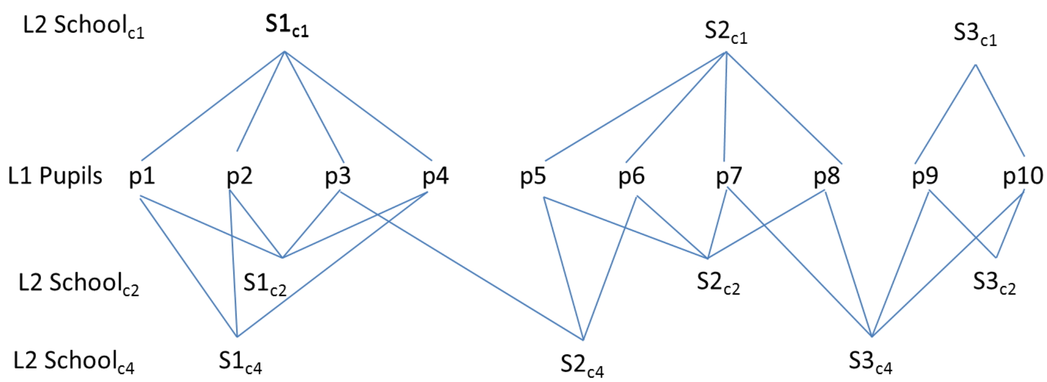
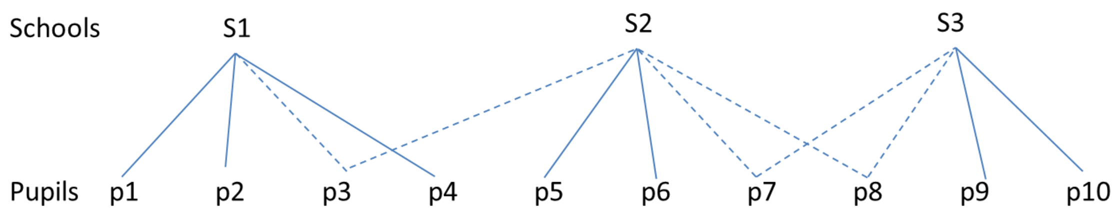


Figure 9: Multiple-Membership Structure representation of Pupils' Mobility



References

Beretvas, S.N. (2010). Cross-classification and multiple membership random effects models. Invited chapter submitted for publication in J.Hox & J.K. Roberts (Eds.) *The handbook of advanced multilevel analysis.*

Beretvas, S.N., & Walter L.Leite (2010). A comparison of Multiple-Membership and Cross-Classification Random Effects Models: Evaluating their Use with Mobile Students' Data. Paper for presentation at the annual meeting of the AERA meeting, Denver.

Harvey Goldstein (2010). *Multiple Statistical Models (4^{th} Edition),* Wiley Series in Probability and Statistics

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks, CA: Sage Publications, Inc.

Rasbash,J.,Browne,W.J.,Goldstein,H.,Yang,M.,Plewis,I.,Healy,M.,et at(2002). *A user's guide to MLwiN (2^{nd} Ed.).* London, UK: Institute of Education

Alastair Leyland & Oyvind Ness (2009). The Effect of area of residence over the life course on subsequent mortality. *Journal of the Royal Statistical Society A (2009) 172, Part3, pp, 555-578*

Antony Fielding & Harvey Goldstein.(2006) Cross-Classification and Multiple Membership Structures in Multilevel Models: An introduction and Review. University of Birmingham 2006

William J Browne, Harvey Goldstein & Jon Rasbash (2001). Multiple membership multiple classification (MMMC) models. *Statistical Modeling* 2001; 1:103-124

Harvey Goldstein, Simon Burgess & Brendon McConnell (2007). Modeling the Effect of Pupil Mobility on School Differences in Educational Achievement. *Journal of the Royal Statistical Society.* Series A, Vol.170, No.4 (2007), pp.941-954

James B. Schreiber & Bryan W. Griffin (2004). Review of Multilevel Modeling and Multilevel Studies in "The Journal of Educational Research" (1992-2002), *The Journal of Educational Research*, Vol.98,No.1, Special Issue on Methodology (2004),pp.24-33

Harvey Goldstein, Jon Rasbash, William Browne, Geoffrey Woodhouse & Michel Poulain (2000). Multilevel Models in the Study of Dynamic Household Structures, *European Journal of Population*, Vol.16, No.4, pp373-387

William Browne (2001). An Introduction to Hierarchical Cross-Classified and multiple membership models, Multilevel Models project Institute of Education, Claritas Senior Analyst Meeting.